doi: 10.1046/j.1529-8817.2003.00066.x

On the Power of Affected Relative Pair Designs for Linkage Studies

Z. LI[1,2,*] and J. L. GASTWIRTH[1,2]

[1]*Department of Statistics, George Washington University, 2201 G Street NW., Washington DC 20052*
[2]*Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, 6120 Executive Blvd., EPS, Bethesda, MD, 20892*

Summary

Using the recurrence risk ratio (λ), Risch (1990b) indicated that affected pairs of distant relatives are preferable to affected sib-pairs for linkage analysis when λ is large and the mode of inheritance is additive. By using the optimum test for affected sib-pairs instead of the test used by Risch (1990b), the range of values of λ for which the affected sib-pair design has larger or smaller power than other pairs is clarified. Risch's conclusion remains true when $\lambda_O > 2.5$, however, sib-pairs have larger power for lower values. As affected sib-pairs occur smore frequently than other relative pairs, when ascertainment costs are non-negligible, they may be the most cost-effective relative pairs to use.

In his seminal article Risch (1990b) demonstrated that affected pairs of distant relatives provide more power in studies of linkage than affected sib-pairs. This report points out that the test used to analyze affected sib pairs was not the optimal one (Knapp, 1994). For affected sib-pair data we show that the means test (Blackwelder & Elston, 1985), the locally optimum one for the additive model considered by Risch, is nearly equivalent to the likelihood ratio test derived from the Neyman-Pearson (NP) theory and has noticeably improved power over the test studied by Risch. For all types of relative pairs, Risch used a test based on n_0 (the number of these pairs having no allele IBD). Except for affected sibs, this procedure is the optimum test because they can only share 0 or 1 allele IBD. In order to compare the power of various statistical designs, one should use the optimum test for each design under consideration. While Risch's conclusion that for the additive mode of inheritance and large genetic effects, e.g. $\lambda_S = \lambda_O \geq 3$, distant relative pairs can provide more power is correct, our results show that for small to modest effects ($\lambda_O \leq 2.5$) the affected sib-pair design is preferable. Here, λ_S is the recurrence risk ratio of disease in siblings and λ_O is the recurrence risk ratio between a parent and offspring (Risch, 1990a,b,c; Rybicki & Elston, 2000). In view of recent interest in complex diseases focusing on discovering genes with modest effects (Cardon & Bell, 2001; Cordell, 2001; Tabor *et al.* 2002), this result indicates that affected sib-pairs, which are often the most convenient relative pairs to ascertain, retain their usefulness.

A brief review of the test used by Risch and the optimal test for affected sib-pair design is helpful. Under the assumption that there is no dominance variance, $V_D = 0$, the sib-pair recurrence risk ratio (λ_S) is the same as that between a parent and offspring (λ_O), i.e., $\lambda_S = \lambda_O$. Following Risch, let n_0, n_1, and n_2 be the number of affected sib-pairs with $IBD = 0, 1, 2$ at the marker locus respectively. The sample size is $N = n_0 + n_1 + n_2$. Thus, n_0, n_1, and n_2 have a trinomial distribution with probabilities:

$$z_{S0} = \frac{1}{4} - \frac{1}{4\lambda_O}(2\Psi - 1)(\lambda_O - 1), \quad z_{S1} = \frac{1}{2},$$

$$z_{S0} = \frac{1}{4} + \frac{1}{4\lambda_O}(2\Psi - 1)(\lambda_O - 1),$$

*Correspondence: Zhaohai Li, Ph.D., George Washington University, Department of Statistics, 315 Funger Hall, Washington, DC 20052. Phone: 202-994-7844. Fax: 202-994-6917. E-mail: zli@gwu.edu

The Power of Various Relative Pair Designs as a Function of λ_O and θ.

θ	Sibs NP	Sibs $n_2 - n_0$	Sibs n_0	Grand P Grand C	Uncle Nephew	Half Sibs	First Cousins
				$\lambda_O = 1.5$			
				$n = 100$			
0.00	0.213	0.225	0.088	0.133	0.133	0.133	0.142
0.05	0.108	0.114	0.043	0.095	0.058	0.068	0.064
0.10	0.052	0.055	0.021	0.066	0.026	0.034	0.028
0.15	0.025	0.026	0.011	0.044	0.012	0.017	0.013
0.20	0.012	0.012	0.006	0.028	0.006	0.009	0.006
0.30	0.003	0.003	0.002	0.011	0.002	0.003	0.002
0.40	0.001	0.001	0.001	0.004	0.001	0.001	0.001
				$n = 200$			
0.00	0.592	0.599	0.334	0.395	0.395	0.395	0.367
0.05	0.338	0.345	0.160	0.290	0.178	0.209	0.168
0.10	0.162	0.167	0.071	0.201	0.071	0.098	0.068
0.15	0.070	0.071	0.031	0.131	0.028	0.043	0.027
0.20	0.028	0.029	0.014	0.080	0.011	0.019	0.011
0.30	0.005	0.005	0.003	0.025	0.003	0.004	0.003
0.40	0.002	0.002	0.001	0.006	0.001	0.001	0.001
				$\lambda_O = 2.5$			
				$n = 100$			
0.00	0.904	0.898	0.705	0.907	0.907	0.907	0.922
0.05	0.631	0.644	0.356	0.797	0.587	0.658	0.666
0.10	0.332	0.348	0.147	0.641	0.261	0.359	0.343
0.15	0.142	0.151	0.057	0.462	0.092	0.156	0.134
0.20	0.054	0.057	0.022	0.296	0.031	0.059	0.046
0.30	0.008	0.008	0.004	0.081	0.004	0.008	0.006
0.40	0.002	0.002	0.001	0.013	0.001	0.002	0.002
				$n = 200$			
0.00	1.000	0.999	0.995	0.999	0.999	0.999	0.999
0.05	0.973	0.97	0.872	0.995	0.952	0.974	0.959
0.10	0.78	0.782	0.522	0.969	0.66	0.794	0.719
0.15	0.431	0.439	0.217	0.886	0.283	0.451	0.349
0.20	0.169	0.173	0.074	0.714	0.088	0.179	0.118
0.30	0.016	0.016	0.008	0.25	0.008	0.017	0.01
0.40	0.002	0.002	0.002	0.03	0.002	0.002	0.002
				$\lambda_O = 5.0$			
				$n = 100$			
0.00	1.000	0.999	0.999	1.000	1.000	1.000	1.000
0.05	0.962	0.953	0.840	1.000	0.991	0.997	0.999
0.10	0.706	0.715	0.430	0.996	0.792	0.903	0.936
0.15	0.353	0.370	0.159	0.963	0.372	0.574	0.628
0.20	0.133	0.141	0.053	0.840	0.117	0.239	0.257
0.30	0.014	0.014	0.007	0.330	0.009	0.021	0.019
0.40	0.002	0.002	0.002	0.038	0.002	0.002	0.002
				$n = 200$			
0.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.05	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.10	0.988	0.985	0.925	1.000	0.994	0.999	0.999
0.15	0.805	0.806	0.553	1.000	0.808	0.947	0.944
0.20	0.407	0.415	0.202	0.997	0.351	0.623	0.593
0.30	0.034	0.035	0.016	0.760	0.021	0.056	0.043
0.40	0.003	0.003	0.002	0.112	0.002	0.003	0.003

where $\Psi = \theta^2 + (1-\theta)^2$. The null hypothesis of no linkage is $H_0 : \theta = \frac{1}{2}$ and the alternative hypothesis is $H_a : \theta = \theta_a$, where $\theta_a < \frac{1}{2}$ is specified. Here, λ_O is a given function of the prevalence of the disease and the additive genetic variance (Risch, 1990a). The test used by Risch (1990b) is: reject H_0 when $n_0 \leq C_\alpha$, where C_α is determined by a prespecified level α. Using the normal approximation, the power of this test is

$$\Phi\left(\frac{\frac{\sqrt{3}}{4}z_\alpha + \sqrt{N}\left(\frac{1}{4} - \mu_a\right)}{\sigma_a}\right),$$

where z_α is αth percentile of the standard normal distribution, Φ is the distribution function of the standard normal distribution, and N is the total number of sib-pairs. The parameters μ_a and σ_a^2 are:

$$\mu_a = \frac{1}{4}(1-b), \quad \sigma_a^2 = \frac{1}{16}(1-b)(3+b),$$
$$b = \frac{1}{\lambda_O}(2\Psi_a - 1)(\lambda_O - 1), \quad \Psi_a = (1-\theta_a)^2 + \theta_a^2.$$

The optimal test based Neyman-Pearson theory is:

$$\text{reject } H_0 \text{ when } n_0 \log(1-b) + n_2 \log(1+b) \geq C_\alpha.$$

The power of the optimal test is given by

$$1 - \Phi\left(\frac{\sqrt{N}(\mu_0 - \mu_a^*)}{\sigma_a^*} + z_{1-\alpha}\frac{\sigma_0}{\sigma_a^*}\right),$$

where

$$\mu_0 = \frac{1}{4}\log(1-b^2), \quad \mu_a^* = \frac{1}{4}[(1-b)\log(1-b) + (1+b)\log(1+b)],$$
$$\sigma_0^2 = \frac{1}{16}\{3[\log(1-b)]^2 + 3[\log(1+b)]^2 - 2[\log(1-b)][\log(1+b)]\},$$
$$\sigma_a^{*2} = \frac{1}{16}\{(1-b)(3+b)[\log(1-b)]^2 + (1+b) \times (3-b)[\log(1+b)]^2 - 2(1-b^2)\log(1-b) \times \log(1+b)\}.$$

When the alternative hypothesis approaches the null hypothesis (i.e., $\theta_a \to \frac{1}{2}$), b goes to zero (i.e., $b \to 0$). Thus, the limiting form of the optimal test becomes: reject H_0 if $n_2 - n_0 \geq C$. Thus, the test statistic $n_2 - n_0 = 2n_2 + n_1 - N$ is locally optimal. Before presenting a numerical comparison of the power of these optimal tests and Risch's for affected sib-pairs, it should be noted that the tests Risch used to analyze data from other relative pairs were optimal.

In the Table we report the power of tests using different relative pairs studied by Risch (1990b) for various recombination fractions (θ). The first three columns report the powers of the three tests for affected sib pairs. The first two are the Neyman-Pearson test and the means test and the third is the test used by Risch. Notice that for all recurrence risk ratios (λ_O) studied the optimal tests for affected sib pairs have greater power than the test based on n_0. Consequently, when $\lambda_O \leq 2.5$ sib-pairs provide greater power than other relative pairs. For larger values of λ_O, Risch's conclusion that studies using distant relative pairs are more powerful remains true even when the optimum test for sib-pairs utilized.

The best choice of an affected relative pair design depends on both the magnitude of the recurrence risk ratio and the probability of ascertaining an affected pair of those relatives. From Risch (1990a), for any type of relative pair (R), the probability that members of a random pair will be affected is $\lambda_R K^2$, where K is the population prevalence rate. Since $\lambda_S > \lambda_R$ for all other types (R) of relative pairs (Risch, 1990a), the probability of ascertaining affected sibs is greater than that of other types of relatives. Thus, when the screening cost is important, the sib-pair design maybe cost-effective. For early onset diseases where it is possible to obtain grandparent-grandchild affected pairs that design has superior power properties for a wide range of θ. This is consistent with recent results of Weinberg (2003) indicating that likelihood tests similar to TDT conditioning on grandparental mating types have substantially greater power compared with tests based on parents.

The above conclusion assumes that conditional on individual's genotype their phenotype is independent of those of other family members. Suppose there is an environmental factor, U, which independently can cause the disease. If closer relatives are more likely to be jointly exposed to U than distant relatives, then the observed recurrence risk for closer relative pairs (sibs) contain a larger non-genetic component than it will for distant relatives. In this situation, a more distant relative pair study may well be preferred. On the other hand, if the factor U is protective, then affected close relatives provide more information about the genetic component.

Acknowledgments

The authors wish to thank Weiping Deng for her assistance with computer program and two referees for helpful comments. This research was partly supported by a grant EY14478 (Z.L.) from NIH and by a grant from NSF (J.L.G.).

References

Blackwelder, W. C. & Elston, R. C. (1985) A comparison of sib-pair linkage tests for disease susceptibility loci. *Genet Epidemiol* **2**, 85–97.

Cardon, L. R. & Bell, J. I. (2001) Association study designs for complex diseases. *Nature Reviews Genetics* **2**, 91–99.

Cordell, H. J. (2001) Sample size requirements to control for stochastic variation in magnitude and location of allele-sharing statistics in affected sibling pairs. *Ann Hum Genet* **65**, 491–502.

Knapp, M., Seuchter, S. A. & Baur, M. P. (1994) Linkage analysis in nuclear families 1: Optimality criteria for affected sib-pair tests. *Hum Hered* **44**, 37–43.

Risch, N. (1990a) Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet* **46**, 222–228.

Risch, N. (1990b) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet* **46**, 229–241.

Risch, N. (1990c) Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *Am J Hum Genet* **46**, 242–253.

Rybicki, B. A. & Elston, R. C. (2000) The relationship between the sibling recurrence- risk ratio and genotype relative risk. *Am J Hum Genet* **66**, 593–604.

Tabor, H. K., Risch, N. & Myers, R. M. (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Reviews Genetics* **3**, 1–7.

Weinberg, C. R. (2003) Studying parents and grandparents to assess genetic contributions to early-onset disease. *Am J Hum Genet* **72**, 438–447.

Received: 15 May 2003
Accepted: 8 August 2003

